

OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/017,736

DATE: 01/02/2002

TIME: 11:32:09

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

4 <110> APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
6 <120> TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
9 <130> FILE REFERENCE: 13/082  
C--> 11 <140> CURRENT APPLICATION NUMBER: US/10/017,736  
C--> 11 <141> CURRENT FILING DATE: 2001-12-14  
11 <150> PRIOR APPLICATION NUMBER: 60/256,031  
12 <151> PRIOR FILING DATE: 2000-12-15  
14 <160> NUMBER OF SEQ ID NOS: 21  
16 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
18 <210> SEQ ID NO: 1  
19 <211> LENGTH: 1230  
20 <212> TYPE: DNA  
21 <213> ORGANISM: HCV  
23 <220> FEATURE:  
24 <221> NAME/KEY: CDS  
25 <222> LOCATION: (1)...(1230)  
27 <400> SEQUENCE: 1

28	atg gac cgg gag atg gct gca tcg tgc gga ggc gcg gtt ttc ata ggt	48
29	Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Ile Gly	
30	1 5 10 15	
32	ctt gca ctc ttg acc ttg tca. cca tac tat aaa gtg ctc ctc gct agg	96
33	Leu Ala Leu Leu Thr Leu Ser Pro Tyr Tyr Lys Val Leu Leu Ala Arg	
34	20 25 30	
36	ctc ata tgg tgg tta cag tat tta atc acc aga gtc gag gcg cac ttg	144
37	Leu Ile Trp Trp Leu Gln Tyr Leu Ile Thr Arg Val Glu Ala His Leu	
38	35 40 45	
40	caa gtg tgg atc ccc cct ctc aat gtt cgg gga ggc cgc gat gcc atc	192
41	Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Ile	
42	50 55 60	
44	atc ctc ctc acg tgc gca gtc cac cca gag cta atc ttt gac atc acc	240
45	Ile Leu Leu Thr Cys Ala Val His Pro Glu Leu Ile Phe Asp Ile Thr	
46	65 70 75 80	
48	aaa ctc ctg ctc gcc ata ttc ggt ccg ctc atg gtg ctc cag gca ggc	288
49	Lys Leu Leu Leu Ala Ile Phe Gly Pro Leu Met Val Leu Gln Ala Gly	
50	85 90 95	
52	ata acc aaa gtg ccg tac ttc gtg cgt gcg cag ggg ctc att cgt gcg	336
53	Ile Thr Lys Val Pro Tyr Phe Val Arg Ala Gln Gly Leu Ile Arg Ala	
54	100 105 110	
56	tgt atg ttg gtg cgg aag gct gcg ggg ggt cat tat gtc caa atg gcc	384
57	Cys Met Leu Val Arg Lys Ala Ala Gly Gly His Tyr Val Gln Met Ala	
58	115 120 125	
60	ttc atg aag cta gct gcg ctg aca ggt acg tac gtt tat gac cat ctc	432
61	Phe Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu	
62	130 135 140	
64	act cca ttg cag gat tgg gcc cac gcg ggc cta cga gac ctt gca gtg	480
65	Thr Pro Leu Gln Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val	
66	145 150 155 160	

ENTERED

## RAW SEQUENCE LISTING

DATE: 01/02/2002

PATENT APPLICATION: US/10/017,736

TIME: 11:32:09

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

```

68 gcg gta gag ccc gtc atc ttc tct gac atg gag gtc aag atc atc acc 528
69 Ala Val Glu Pro Val Ile Phe Ser Asp Met Glu Val Lys Ile Ile Thr
70                               165                               170                               175
72 tgg ggg gcg gac acc gcg gca tgc ggg gac atc att tca ggt ctg ccc 576
73 Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Ser Gly Leu Pro
74                               180                               185                               190
76 gtc tcc gct cga agg gga agg gag ata ctc ctg gga ccg gcc gat aat 624
77 Val Ser Ala Arg Arg Gly Arg Glu Ile Leu Leu Gly Pro Ala Asp Asn
78                               195                               200                               205
80 ttt gaa ggg cag ggg tgg cga ctc ctt gcg ccc atc acg gcc tac tcc 672
81 Phe Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser
82                               210                               215                               220
84 caa cag aca cgg ggc cta ctt ggt tgc atc atc acc agc ctc aca ggc 720
85 Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly
86 225                               230                               235                               240
88 cgg gac aag aac cag gtc gag ggg gag gtt caa gtg gtc tcc acc gct 768
89 Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala
90                               245                               250                               255
92 aca caa tct ttc ctg gcg acc tgc gtc aac ggc gtg tgt tgg act gtc 816
93 Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val
94                               260                               265                               270
96 ttc cat ggc gcc ggc tca aag acc ttg gcc ggc ccc aaa ggc cca atc 864
97 Phe His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile
98                               275                               280                               285
100 acc cag atg tac act aat gtg gac cag gac ctc gtc ggc tgg cag gcg 912
101 Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala
102                               290                               295                               300
104 ccc cct ggg gcg cgc tcc atg aca cca tgc acc tgc ggc agc tcg gac 960
105 Pro Pro Gly Ala Arg Ser Met Thr Pro Cys Thr Cys Gly Ser Ser Asp
106 305                               310                               315                               320
108 ctc tat ttg gtc acg aga cat gcc gac gtc att ccg gtg cgc cgg cgg 1008
109 Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg
110                               325                               330                               335
112 ggc gac agt agg ggg agc ctg ctc tcc ccc agg cct gtc tcc tac ttg 1056
113 Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu
114                               340                               345                               350
116 aag ggc tct tcg ggt ggc cca ctg ctc tgc cct tcg ggg cac gct gtg 1104
117 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val
118                               355                               360                               365
120 ggc atc ttc cgg gct gct gtg tgc acc cgg ggg gtt gca aaa gcg gtg 1152
121 Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val
122                               370                               375                               380
124 gac ttc ata cct gtt gag tct atg gaa act acc atg cgg act agt agc 1200
125 Asp Phe Ile Pro Val Glu Ser Met Glu Thr Thr Met Arg Thr Ser Ser
126 385                               390                               395                               400
128 gct tgg cgt cac ccg cag ttc ggt ggt taa 1230
129 Ala Trp Arg His Pro Gln Phe Gly Gly *
130                               405
133 <210> SEQ ID NO: 2

```

## RAW SEQUENCE LISTING

DATE: 01/02/2002

PATENT APPLICATION: US/10/017,736

TIME: 11:32:09

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

134 &lt;211&gt; LENGTH: 409

135 &lt;212&gt; TYPE: PRT

136 &lt;213&gt; ORGANISM: HCV

138 &lt;400&gt; SEQUENCE: 2

```

139 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Ile Gly
140 1      5      10      15
141 Leu Ala Leu Leu Thr Leu Ser Pro Tyr Tyr Lys Val Leu Leu Ala Arg
142      20      25      30
143 Leu Ile Trp Trp Leu Gln Tyr Leu Ile Thr Arg Val Glu Ala His Leu
144      35      40      45
145 Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Ile
146      50      55      60
147 Ile Leu Leu Thr Cys Ala Val His Pro Glu Leu Ile Phe Asp Ile Thr
148 65      70      75      80
149 Lys Leu Leu Leu Ala Ile Phe Gly Pro Leu Met Val Leu Gln Ala Gly
150      85      90      95
151 Ile Thr Lys Val Pro Tyr Phe Val Arg Ala Gln Gly Leu Ile Arg Ala
152      100     105     110
153 Cys Met Leu Val Arg Lys Ala Ala Gly Gly His Tyr Val Gln Met Ala
154      115     120     125
155 Phe Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu
156      130     135     140
157 Thr Pro Leu Gln Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val
158 145     150     155     160
159 Ala Val Glu Pro Val Ile Phe Ser Asp Met Glu Val Lys Ile Ile Thr
160      165     170     175
161 Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Ser Gly Leu Pro
162      180     185     190
163 Val Ser Ala Arg Arg Gly Arg Glu Ile Leu Leu Gly Pro Ala Asp Asn
164      195     200     205
165 Phe Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser
166      210     215     220
167 Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly
168 225     230     235     240
169 Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala
170      245     250     255
171 Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val
172      260     265     270
173 Phe His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile
174      275     280     285
175 Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala
176      290     295     300
177 Pro Pro Gly Ala Arg Ser Met Thr Pro Cys Thr Cys Gly Ser Ser Asp
178 305     310     315     320
179 Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg
180      325     330     335
181 Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu
182      340     345     350
183 Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/017,736

DATE: 01/02/2002

TIME: 11:32:09

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

```

184          355          360          365
185 Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val
186          370          375          380
187 Asp Phe Ile Pro Val Glu Ser Met Glu Thr Thr Met Arg Thr Ser Ser
188 385          390          395          400
189 Ala Trp Arg His Pro Gln Phe Gly Gly
190          405
193 <210> SEQ ID NO: 3
194 <211> LENGTH: 1011
195 <212> TYPE: DNA
196 <213> ORGANISM: HCV
198 <220> FEATURE:
199 <221> NAME/KEY: CDS
200 <222> LOCATION: (1)...(1005)
202 <400> SEQUENCE: 3
203 atg aaa aag aaa aag ctc gag cat cac cat cac cat cac act agt gca 48
204 Met Lys Lys Lys Lys Leu Glu His His His His His His Thr Ser Ala
205 1          5          10          15
207 ggc ata acc aaa gtg ccg tac ttc gtg cgt gcg cag ggg ctc att cgt 96
208 Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala Gln Gly Leu Ile Arg
209          20          25          30
211 gcg tgt atg ttg gtg cgg aag gct gcg ggg ggt cat tat gtc caa atg 144
212 Ala Cys Met Leu Val Arg Lys Ala Ala Gly Gly His Tyr Val Gln Met
213          35          40          45
215 gcc ttc atg aag cta gct gcg ctg aca ggt acg tac gtt tat gac cat 192
216 Ala Phe Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His
217          50          55          60
219 ctc act cca ttg cag gat tgg gcc cac gcg ggc cta cga gac ctt gca 240
220 Leu Thr Pro Leu Gln Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala
221 65          70          75          80
223 gtg gcg gta gag ccc gtc atc ttc tct gac atg gag gtc aag atc atc 288
224 Val Ala Val Glu Pro Val Ile Phe Ser Asp Met Glu Val Lys Ile Ile
225          85          90          95
227 acc tgg ggg gcg gac acc gcg gca tgc ggg gac atc att tca ggt ctg 336
228 Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Ser Gly Leu
229          100          105          110
231 ccc gtc tcc gct cga agg gga agg gag ata ctc ctg gga ccg gcc gat 384
232 Pro Val Ser Ala Arg Arg Gly Arg Glu Ile Leu Leu Gly Pro Ala Asp
233          115          120          125
235 aat ttt gaa ggg cag ggg tgg cga ctc ctt gcg ccc atc acg gcc tac 432
236 Asn Phe Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr
237          130          135          140
239 tcc caa cag aca cgg ggc cta ctt ggt tgc atc atc acc agc ctc aca 480
240 Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr
241 145          150          155          160
243 ggc cgg gac aag aac cag gtc gag ggg gag gtt caa gtg gtc tcc acc 528
244 Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr
245          165          170          175
247 gct aca caa tct ttc ctg gcg acc tgc gtc aac ggc gtg tgt tgg act 576

```

## RAW SEQUENCE LISTING

DATE: 01/02/2002

PATENT APPLICATION: US/10/017,736

TIME: 11:32:09

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

```

248 Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr
249      180      185      190
251 gtc ttc cat ggc gcc ggc tca aag acc ttg gcc ggc ccc aaa ggc cca 624
252 Val Phe His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro
253      195      200      205
255 atc acc cag atg tac act aat gtg gac cag gac ctc gtc ggc tgg cag 672
256 Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln
257      210      215      220
259 gcg ccc cct ggg gcg cgc tcc atg aca cca tgc acc tgc ggc agc tcg 720
260 Ala Pro Pro Gly Ala Arg Ser Met Thr Pro Cys Thr Cys Gly Ser Ser
261 225      230      235      240
263 gac ctc tat ttg gtc acg aga cat gcc gac gtc att ccg gtg cgc cgg 768
264 Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg
265      245      250      255
267 cgg ggc gac agt agg ggg agc ctg ctc tcc ccc agg cct gtc tcc tac 816
268 Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr
269      260      265      270
271 ttg aag ggc tct tcg ggt ggc cca ctg ctc tgc cct tcg ggg cac gct 864
272 Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala
273      275      280      285
275 gtg ggc atc ttc cgg gct gct gtg tgc acc cgg ggg gtt gca aaa gcg 912
276 Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala
277      290      295      300
279 gtg gac ttc ata cct gtt gag tct atg gaa act acc atg cgg act agt 960
280 Val Asp Phe Ile Pro Val Glu Ser Met Glu Thr Thr Met Arg Thr Ser
281 305      310      315      320
283 agc gct tgg cgt cac ccg cag ttc ggt ggt aaa aag aaa aag taa 1005
284 Ser Ala Trp Arg His Pro Gln Phe Gly Gly Lys Lys Lys Lys *
285      325      330
287 ggatcc 1011
289 <210> SEQ ID NO: 4
290 <211> LENGTH: 334
291 <212> TYPE: PRT
292 <213> ORGANISM: HCV
294 <400> SEQUENCE: 4
295 Met Lys Lys Lys Lys Leu Glu His His His His His His Thr Ser Ala
296 1      5      10      15
297 Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala Gln Gly Leu Ile Arg
298      20      25      30
299 Ala Cys Met Leu Val Arg Lys Ala Ala Gly Gly His Tyr Val Gln Met
300      35      40      45
301 Ala Phe Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His
302      50      55      60
303 Leu Thr Pro Leu Gln Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala
304 65      70      75      80
305 Val Ala Val Glu Pro Val Ile Phe Ser Asp Met Glu Val Lys Ile Ile
306      85      90      95
307 Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Ser Gly Leu
308      100      105      110

```

Use of n and/or Xaa has been detected in the Sequence Listing.  
 Review the Sequence Listing to insure a corresponding  
 explanation is presented in the <220> to <223> fields of  
 each sequence using n or Xaa.

## VERIFICATION SUMMARY

DATE: 01/02/2002

PATENT APPLICATION: US/10/017,736

TIME: 11:32:10

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\01022002\J017736.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:851 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:19

L:870 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20